****

A Plain White Paper

**Fraud Detection Using Amazon Web Services**

**-- Pradyumn Nukala --**

FRAUD DETECTION USING AMAZON WEB SERVICES

A PLAIN WHITE PAPER

Introduction

Fraud is a billion-dollar business and it’s increasing every year. I a PwC survey conducted in 2018, 49% of the companies asked had dealt with fraud. This is an increase from the 2016 study where 36% of companies were affected. Fraud detection is an ongoing problem that can cost a business a large amount of money while also damaging customer.

Traditional methods of data analysis have been used to detect fraud for a long time. They require complex and time-consuming investigations that deal with different domains of knowledge like finance, economics, business practices, and law. Fraud often consists of many instances or incidents involving repeated transgressions using the same method. Fraud instances can be similar in content and appearance but usually not identical.

A popular method for detecting fraud currently are rule-based engines. There are several problems with these methods which include that they require active management and are a one size fits all solution. The information silos in these engines ignore the benefits of collective intelligence, this leads to the rules being prone to human errors and bias.

Machine Learning as the Solution

Machine Learning provides a much more flexible approach to address constantly evolving Fraud Tactics, providing a better solution than preexisting methods that utilize rigid active rule-based systems.

• Wide Parameters: Machine learning and artificial intelligence solutions may be classified into two categories: supervised and unsupervised learning. These methods seek for accounts, customers, suppliers that behave unusually in order to output suspicion scores, rules, or visual anomalies depending on the method.

• Limitations: Despite the methods used the output gives us only an indication of fraud likelihood. No stand-alone statistical analysis can assure that a particular transaction is a fraudulent one, but they can identify them to high degrees of accuracy.

Use Cases

The market for fraudulent transactions or data is relevant to almost every major industry. Implementing fraud detection in these use cases are ones that I found:

**• Insurance Fraud**: Firms are now utilizing internal data such as customer social media, call center notes, personal details, and voice recordings to gain insight into potentially fraudulent claims. I found, a claimant may declare their car was damaged in flooding, but their social media shows they were at a completely different location. The required dataset to pursue this implementation would need claim details, and aggregated details about the customer in question.

**• Telecom Fraud**: Firms in the telecommunication industry analyze call record data coupled with customer data in order to build a complete profile. Fraud in this industry is generally demonstrated when customers don’t pay bills and have suspicious long-distance calling patterns within six months of joining the service. The dataset required for this solution would need to have call record info and a general customer profile.

Testing

Before implementing a model, I had tested various algorithms locally using SKLearn and the dataset provided for by Amazon. Utilizing seven different algorithms to find which I got the following scores for:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Classification Trees** | **KNN** | **Linear Discriminant** | **Logistic Regression** | **Random Forest** | **XGBoost** |
| **Sensitivity** | 0.74 | 0.67 | 0.72 | 0.58 | 0.53 | 0.53 |
| **F1 Score** | 0.73 | 0.76 | 0.77 | 0.68 | 0.65 | 0.65 |

Based on these results we can see that the two best algorithms for having the highest accuracy and create innate business value would be the **Random Forest** and **XGBoost** Algorithms. It would be more logical to utilize the **Random Forest** Algorithm as it yields similar prediction capability at lower compute power. From my research I have found that in identifying fraud the most important metric would be **Sensitivity** followed by **F1 Score**. Sensitivity Measures Identifying Fraud. F1 Score Measures Legitimate Customers.

Implementation

The solution created was made using Amazon Web Services as the backend, Implementing the SageMaker solution to create an endpoint.

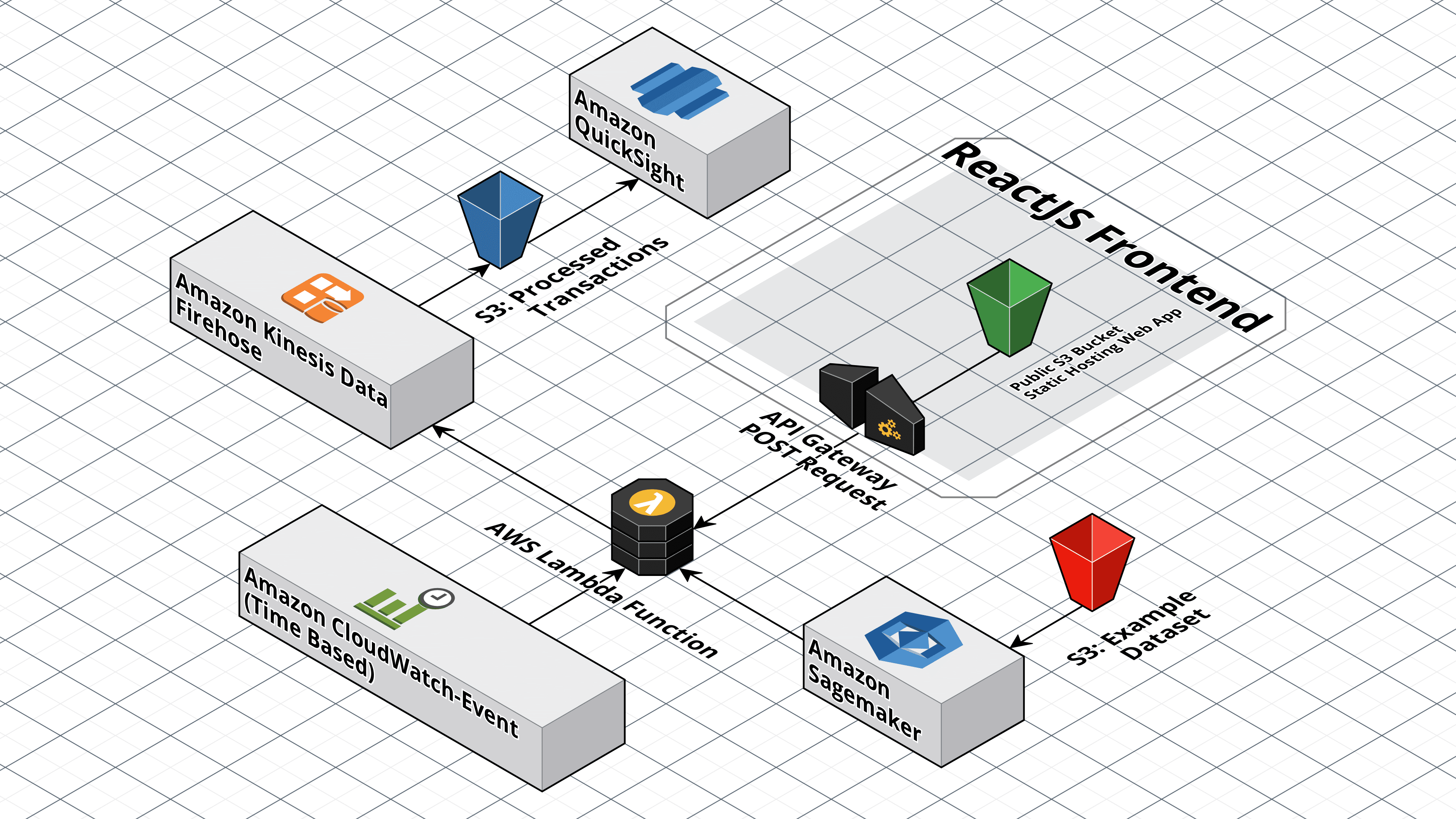
**• CloudWatch**: This event is configured to run every minute and triggers a Lambda Function that processes transactions from the example dataset.

**• Lambda Function**: The lambda function is set to process transactions from the dataset and invokes the SageMaker endpoint

**• SageMaker**: Utilizing the Linear Learner model predicts whether transactions are fraudulent.

**• Kinesis**: The delivery stream loads the processed transactions into another Amazon S3 bucket for storage.

**• QuickSight**: Pulls the transactions from S3, for visualization, reporting, ad-hoc queries, and more detailed analysis.

In the frontend of this solution, a public S3 bucket hosts a static webpage which calls a Lambda Function on AWS. The request uses API Gateway to call onto the SageMaker which returns a prediction in JSON format to the frontend.

Conclusion

A couple of the biggest challenges with preexisting rule-based systems follow:

1. Rules are effective only when they are actively monitored and managed by dedicated fraud teams. This means fraud teams need to be staffed to review manual review queues, rejects and chargebacks and recalculate thresholds that can be codified in a rule. This makes fraud prevention reactive in a fast-moving business environment.
2. Rules are trained to do what is told without adding intelligence. They follow a binary view of whether the rules criteria are met or not. They do not dynamically adjust themselves for normal behavior or seasonal fluctuations. This would result in more false positives and unhappy customers.
3. Rules that are based on a single channel/device do not provide a holistic view of consumer activity across multiple channels. 68% of fraud today is a cross channel. Thus, the importance of building a complete customer profile across channels.
4. Incorrect and poorly coded rules increase manual review queues and continue to result in high fraud rates.

The implementation created can be scaled for any fraud detection system outside of credit card transactions. This could be made by preprocessing a dataset and utilizing the created model. After this is done there would be little reconfiguration need to be made on the frontend side.

About Unissant

Unissant is an advanced data analytics and business transformation services provider with expertise in healthcare and health IT, finance, national security, and energy. The company delivers innovative solutions to assist government agencies and private sector businesses in tackling their biggest challenges. Founded in 2006, Unissant is a prime contractor on various government vehicles such as CIO-SP3, GSA PSS, GSA HealthIT SIN, and GSA 8(a) STARS II and is a CMMI Level 3, ISO 9001 & 27001 certified company headquartered in Herndon, Virginia with a satellite office in San Antonio, Texas. In March 2017, Unissant received the Government Project of the Year award by Small and Emerging Contractors Advisory Forum (SECAF)

# COPYRIGHT

Copyright © 2018 Unissant, Inc. All Rights Reserved.

Restricted Rights Legend

This document may not, in whole or in part, be copied photocopied, reproduced, translated, or reduced to any electronic medium or machine-readable form without prior consent, in writing, from Unissant, Inc.

Information in this document is subject to change without notice and does not represent a commitment on the part of Unissant. This document is provided “as is” without warranty of any kind including without limitation, any warranty of merchantability or fitness for a particular purpose. Further, Unissant does not warrant, guarantee, or make any representations regarding the use, or the results of the use, of the written material in terms of correctness, accuracy, reliability, or otherwise.

All other brand and product names are trademarks of their respective companies.

For additional information on Unissant full range of services, please visit our website at [**www.unissant.com**](http://www.unissant.com) or call us at **703.889.8500.**